Encryption Struggles Persist: When Tech-Savvy Students Face Challenges with PGP in Thunderbird

1st Md Imanul Huq
Department of CSE
Texas A&M University
College Station, USA
imanulhuq@tamu.edu

2nd Ahmed Tanvir Mahdad
Department of CSE
Texas A&M University
College Station, USA
mahdad@tamu.edu

3rd Nitesh Saxena
Department of CSE
Texas A&M University
College Station, USA
nsaxena@tamu.edu

***Abstract*—This longitudinal study explores recurring usability challenges faced by students in a university-level cybersecurity course during their first use of Thunderbird for PGP (Pretty Good Privacy) email encryption. Despite being tech-savvy and security-aware, students encountered persistent issues such as public key import failures, unintuitive interfaces, and a lack of feedback on successful encryption.**

Our multifaceted analysis included sentiment classification via the Hugging Face transformer pipeline, Google Trends to assess global search behavior, and N-gram/word cloud visualizations of student support emails. We also examined correlations between Thunderbird version updates and support inquiries, revealing a strong relationship between software changes and usability friction.

That even technically proficient users struggled highlights a critical concern: if Thunderbird's PGP features hinder advanced users, general users are likely to face even greater barriers. These findings underscore the urgent need for user-centered improvements in Thunderbird's PGP integration—particularly in UI clarity, cross-platform consistency, and feedback mechanisms.

I. Introduction

Pretty Good Privacy (PGP) is a widely adopted protocol for securing email communication, ensuring both message confidentiality and authenticity [1]. It encrypts messages using the recipient's public key and decrypts them with the sender's private key, making it vital for safeguarding sensitive exchanges. Integrating PGP into cybersecurity education equips students with essential skills—but using it in practice often reveals usability barriers.

PGP's usability challenges are well documented. Whitten and Tygar's foundational study showed that most participants failed to encrypt messages in PGP 5.0 due to complex interfaces and poor user feedback [2]. Follow-up studies also highlight difficulties in key management, unintuitive GUI designs, and user misunderstandings—even among motivated users [14], [15], [19], [20], [23].

Thunderbird, Mozilla's open-source email client, now offers native PGP support (since version 78), aiming to streamline encryption by removing the need for third-party plugins like Enigmail [12]. However, significant usability issues remain—especially around key imports, cross-platform behavior, and unclear encryption feedback [3].

This work was partially supported by grants NSF CNS-2201465, CNS-2154507 and OAC-2139358.

In this paper, we investigate these issues through a multi-semester, real-world usability study involving students in a university-level cybersecurity course (Fall 2022–Fall 2024). Despite detailed instructions and high technical proficiency, students struggled with public key import errors, UI inconsistencies (notably on macOS), and an absence of confirmation cues for successful encryption. These repeated struggles reveal important shortcomings in Thunderbird's PGP implementation.

To uncover deeper patterns, we conducted a multifaceted analysis, including sentiment classification using Hugging Face transformers, Google search trend mining, N-gram/word cloud visualizations of student emails, and correlation studies linking Thunderbird version updates to support inquiries. Our findings illuminate persistent usability pain points and opportunities for improvement.

Understanding how tech-savvy users experience PGP tools like Thunderbird is crucial for improving accessibility for broader populations. Enhancing usability not only benefits students but also supports professionals in journalism, law, and activism—fields where encrypted communication is critical.

This paper makes the following contributions:

- **Real-world usability study:** We analyze real classroom experiences across four semesters, capturing longitudinal friction points and response patterns.
- **Multi-modal empirical analysis:** Our methodology integrates support ticket review, TA-hour tracking, NLP-based sentiment and phrase mining, and global search behavior analysis to triangulate user pain points.
- **Correlation of updates with usability:** We reveal a strong correlation between Thunderbird version releases and increased user confusion, using Pearson coefficient analysis and visual metrics.
- **Actionable design recommendations:** Based on student struggles, we propose practical UI/UX improvements for Thunderbird, including clearer feedback, smarter key import logic, and embedded tooltips.

Unlike prior studies that rely on controlled experiments or surveys, our research is grounded in authentic user behavior within an educational setting. This longitudinal, observational

approach captures not just initial barriers, but how usability problems persist even after repeated instructional refinement.

Broader Relevance Beyond Thunderbird: Though focused on Thunderbird, the issues we identify—unintuitive key management, lack of feedback, and version-related incompatibility—mirror usability barriers across other PGP tools like Enigmail, ProtonMail, and GPG-based clients. As such, our findings offer generalizable insights into the persistent friction users face when adopting public key cryptography, providing guidance for developers and educators alike.

II. Background

OpenPGP in Thunderbird enables users to encrypt and decrypt emails directly within the email client. A public key, which is freely distributed, is used for encryption, while a private key, kept confidential, is used for decryption. This ensures that only the intended recipient can decrypt and read the encrypted messages.

Numerous email clients are available in the current market [28]. According to Litmus, the leading clients include Apple Mail, Gmail, Outlook, Yahoo Mail, Thunderbird, GMX, and ProtonMail. For our study, we required an email client that inherently supports PGP encryption. As illustrated in Table I, only Thunderbird and ProtonMail offer built-in PGP encryption capabilities. However, due to ProtonMail's limited market share of only 0.17%, as reported by 6sense [29], we selected Thunderbird for our analysis.

OpenPGP is a widely used, non-proprietary protocol for securing data communications through encryption and authentication, as outlined in RFC 4880 [31]. Developed by Phil Zimmermann in 1991, OpenPGP ensures that cryptographic elements like public keys can be used seamlessly across various platforms due to its standardized and platform-agnostic design. This interoperability is achieved through specific formats and encoding methods, such as the standardized structure of public-key packets and the use of big-endian format for scalar numbers. These design principles make OpenPGP a robust framework for maintaining secure communications across different operating systems and applications, reflecting Zimmermann's vision of democratizing access to strong encryption [31].

However, we found that neither ProtonMail nor Thunderbird allows users to import public keys for PGP encryption vice versa. Therefore, we selected Thunderbird for its popularity.

TABLE I: Email Clients with PGP Support

Client	PGP Support	Remarks
Thunderbird	Yes	Native support since v-78.
Outlook	No	Needs Gpg4win & GpgOL plugin.
Apple Mail	No	Needs GPGTools.
Gmail	No	Needs Mailvelope extension.
ProtonMail	Yes	Built-in support for PGP.
Yahoo Mail	No	Needs Mailvelope extension.

Public key cryptography forms the foundation of PGP encryption. Each user possesses a key pair consisting of a public key and a private key. The public key is used by others to encrypt messages intended for the user, while the private key is used to decrypt these messages, ensuring that only the intended recipient can access the message content. Thunderbird facilitates the generation of new key pairs or the importation of existing ones, making it a practical tool for email encryption.

The course covers a comprehensive study of network communication vulnerabilities and their corresponding defense mechanisms, with a strong emphasis on cryptographic algorithms and their real-world applications. The hands-on exercises, including the use of PGP encryption in Thunderbird, are designed to provide students with practical experience in securing email communications.

Addressing these usability challenges is crucial not only for students but for all types of stakeholders who rely on secure email communication. For instance, journalists who need to protect their sources, activists who require anonymity to safeguard their identities, and professionals handling sensitive information all depend on effective and user-friendly encryption tools. By improving Thunderbird's PGP integration, we can enhance the overall user experience, making the process of securing email communications more intuitive and accessible for everyone

III. Epistemic Environment

Our experience with Thunderbird's issues arose during a university-level course exercise conducted asynchronously with recorded lectures delivered through the Canvas platform [27]. This course covers networking communication vulnerabilities and corresponding defense mechanisms, starting with cryptographic algorithms and their application in real-world network security. It addresses vulnerabilities and designs defense mechanisms across various layers, including application (e.g., PGP email), transport (SSL/TLS), network (IPSec), and data link (WEP/WPA) security. The Thunderbird exercise was introduced in Module 1.

A. Course Learning Outcome

Students will learn and analyze cryptographic algorithms, apply them to real-world network security applications, explore cryptographic techniques, identify and analyze networking vulnerabilities, and design effective defense mechanisms to ensure robust security. The course also covers secure communication protocols, such as PGP, SSL/TLS, VPNs, and WEP/WPA, equipping students with the skills needed to enhance network security across various layers and platforms.

B. The Thunderbird Exercise

As discussed previously, the hands-on exercise was designed to familiarize students with PGP email encryption using Thunderbird, emphasizing practical cryptographic principles crucial for a computer and network security course [3]. Students installed Thunderbird on their devices using personal email accounts for smooth setup.

They imported a public key from Canvas into Thunderbird's "OpenPGP Key Manager" to communicate securely with the TA, configuring encryption settings and sending a signed, encrypted email to demonstrate their proficiency. Reading and decrypting messages further reinforced their understanding

using Thunderbird's decryption capabilities [4]. This exercise effectively bridged theoretical knowledge with practical skills in network security.

IV. Methodology

This research investigates the challenges faced by students using Thunderbird for PGP email encryption in a university-level cybersecurity course. Based on our observations across semesters, we removed any personally identifiable information during analysis. Since the study relies on observations rather than direct data collection from students, ethical considerations such as informed consent and anonymity are not applicable. The methodology employed here involves a qualitative approach:

Coursework Observations: The primary source of data comes from observing student interactions with Thunderbird and PGP during a specific homework assignment in the course. This assignment required students to send PGP-encrypted emails using the provided public key and Thunderbird.

Student Inquiries and Support Tickets: Throughout the semester, the author, as a TA, provided support to students encountering difficulties with PGP and Thunderbird. Analyzing the nature of these inquiries and support tickets offered insights into the specific problems students faced.

The data collection process involved: We, as instructors and TAs of the course, undertook a comprehensive analysis of the student challenges with Thunderbird and PGP. Throughout the semesters, we maintained detailed notes documenting student challenges with Thunderbird and PGP. These notes included descriptions of the issues, any error messages encountered, and the students' operating systems (e.g., Windows, Mac).

Text Preprocessing for Sentiment Analysis

To ensure accurate and unbiased sentiment analysis, we implemented a structured text preprocessing pipeline tailored for student support emails. The goal was twofold: to preserve the semantic content necessary for sentiment detection and to eliminate any elements that could introduce noise or ethical concerns, such as personal identifiers. This step was critical because irrelevant or overly personalized content can distort model predictions or violate student privacy norms, even in observational research.

Our preprocessing pipeline included the following stages:

- **Named Entity Removal:** Using spaCy's Named Entity Recognition (NER), we identified and removed all personal entities to protect student anonymity while preventing sentiment misclassifications based on names.
- **UIN and Email Address Removal:** We applied regular expressions to strip University Identification Numbers and email addresses. These identifiers are not relevant for sentiment and could bias the model due to their structured format.
- **Greeting and Sign-off Removal:** Standard opening and closing phrases (e.g., "Hi," "Dear Professor," "Thanks") were eliminated using regex patterns. Such formalities often express politeness rather than sentiment, and their inclusion could falsely inflate positivity scores.
- **Sign-off with Name Cleanup:** We applied additional regex filters to catch common name-containing closings that were not detected by NER (e.g., "Thanks, John"). This ensured comprehensive anonymization and consistent tone assessment.
- **Whitespace and Formatting Normalization:** Excessive line breaks, indentation, and extra spaces were removed to ensure clean, uniform input for the model. This helped improve the consistency of tokenization and reduced the risk of unpredictable parsing behavior.

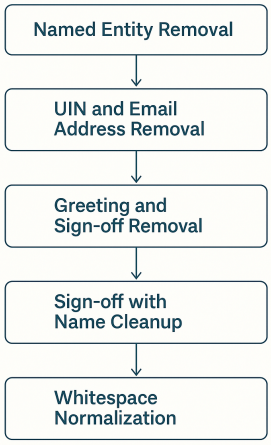

Fig. 1: Text Preprocessing Pipeline for Student Email Sentiment Analysis

Through this preprocessing, each email was distilled down to its core message—the portion most reflective of the student's emotional and cognitive experience with Thunderbird and PGP. These refined inputs were then passed to the Hugging Face sentiment analysis pipeline [26], enabling high-confidence classification without compromising ethical or technical integrity.

Email and TA Hour Analysis: We conducted a thorough review of student support tickets related to PGP and Thunderbird during the observed semesters. These tickets provided valuable insights into the specific issues students faced. This qualitative data was analyzed thematically, revealing common recurring challenges in public key management within Thunderbird. These themes underpin the findings presented in subsequent sections.

Ethical Considerations: This study was based on observational data gathered in the natural flow of coursework support, where no interventions, experiments, or direct student recruitment occurred. The analysis used de-identified student support emails and Zoom logs that contained no personally identifiable information. All data were sanitized using automated NLP techniques to remove names, emails, UINs, and sign-offs, ensuring privacy and anonymity. Since this work falls under educational process improvement and uses secondary anonymized data, it did not require formal Institutional Review Board (IRB) oversight as per institutional policy. Nonetheless, we adhered to ethical best practices in handling and reporting all student-related content, and no data were used beyond what was organically produced during course participation.

V. Findings

We list our findings based on observations across the semesters, combined with our analysis of student support tickets. These findings highlight several recurring issues and potential areas for improvement in Thunderbird's PGP implementation. In table II, we have tabulated the frequency of common issue.

TABLE II: Summary of Issues in Thunderbird

Category	Frequency
Version Mismatch	High
Unintuitive GUI	Medium
Lack of Feedback	Medium
Using University Email	Low

1. Version Mismatch: Errors importing public keys due to incompatibility between Thunderbird version and key format. This is the most prevalent issue observed was public key import errors. This often stemmed from incompatibility between the version of Thunderbird students were using and the format of the provided public key. Students might have been using older keys generated with earlier Thunderbird versions, leading to import failures. This will be discussed further.

2. Unintuitive GUI: Several students, particularly those on Mac, expressed confusion regarding the user interface for managing PGP keys within Thunderbird. The lack of intuitive design elements has contributed to difficulties in navigating key import and encryption processes. For example, in macOS, students had to access the Apple app menu to access end-to-end encryption in Thunderbird, while this option was available directly in the app interface on Microsoft. This discrepancy

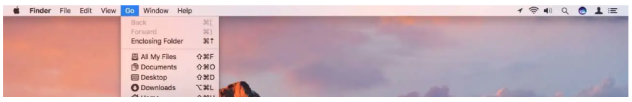

Fig. 2: The user interface for managing PGP keys in Thunderbird on macOS, requiring access through the Apple app menu.

puzzled both our students and the TA. One student expressed their confusion:

"I believe my OpenPGP key manager does not have the import button? Am I right in this conjecture, or am I missing something?"

This misunderstanding occurred because the student attempted to locate the import key option within the Thunderbird, unaware that it was actually accessed through the Apple app menu.

3. Lack of Feedback: Students emailed uncertainty about whether their emails were successfully encrypted. The absence of clear confirmation or error notifications within Thunderbird created confusion about the effectiveness of their PGP implementation. TA had to send separate emails confirming if their encryption was correct. While this approach does not fully address the purpose of PGP, we believe it could help users.

4. Extra Line in Key: A less frequent issue involved public keys containing an extra line not recognized by Thunderbird. This could potentially be caused by errors during key generation or improper formatting. We can see that for the public key in the following figure, there is an empty line between the beginning of the key and the code. However, with a different version of

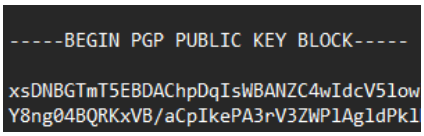

Fig. 3: Extra line in PGP Public Key

Thunderbird, we did not find any line while generating public. This caused a problem importing public keys. After struggling

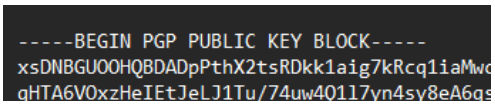

Fig. 4: No extra line in PGP Public Key

for a long time, a student found a solution and wrote:

"I figured it out – the error was that there was an extra newline at the top of the key. I removed this and it imported successfully!"

This information was quite perplexing to us. This is unacceptable just because of one empty line whole functionality breaks down.

5. Using University Email: A small number of students initially attempted to use their university email account for PGP encryption, which is not recommended due to potential security limitations. One student mentioned:

".....It is working if I use my personal Gmail account....."

This might be counterproductive as we can assume a big portion of users of PGP encryption would be from security-sensitive areas. And a big portion of this demographic would use an institution-affiliated email address. These findings indicate that Thunderbird's public key management poses challenges for students unfamiliar with PGP. Version incompatibility, an unintuitive interface, and a lack of clear feedback contribute to their frustration and hinder the learning experience. Many students focus on quickly completing the exercise, which can lead to real-world problems. Despite our testing and instructions, students often fail to use the specified Thunderbird version. For instance, if Alice wants to send Bob an encrypted PGP email, she might not know which Thunderbird version he uses. Additionally, Alice may struggle with the macOS interface. We provided comprehensive, step-by-step instructions, yet some users still installed the wrong version.

From table III we can see in Fall 22, 7.32% of graduate students (6 out of 82) sent emails about Thunderbird's PGP issues, particularly Public Key Import failures. In Fall 23, 20% of undergraduates (14 out of 70) and 21.88% of graduates (14 out of 64) reported similar problems. In Spring 24, 20% of undergraduates (10 out of 50) and 12.5% of graduates

(10 out of 80) faced the same issues. Despite their technical proficiency and our detailed instructions, these students struggled with Thunderbird's PGP functionality, highlighting the need for more intuitive and user-friendly tools. All students eventually managed to send emails using encryption, but only after overcoming these difficulties. It is to be noted that in Fall 22 there were no undergrad students.

TABLE III: Percentage of Students Who Reported Issues with Email

Semester	Student Type	Emails Sent	Percentage (%)
Fall 22	Undergrad	0/0	N/A
	Grad	6/82	7.32
Fall 23	Undergrad	14/70	20
	Grad	14/64	21.88
Spring 24	Undergrad	10/50	20
	Grad	10/80	12.5

Due to the limited data points, traditional statistical tests such as t-tests or ANOVA are not applicable in this case. Each semester's data consists of only one value per group, which is insufficient for these tests. Instead, the analysis focuses on observed trends and percentages: For graduate students, the percentage of students emailing about issues peaked at 21.88% in Fall 23, decreasing to 12.50% in Spring 24. Compared to Fall 22 (7.32%), there was a notable increase in Fall 23 followed by a decrease in Spring 24. For undergraduate students, the issue of public key import in Thunderbird remained stable, with 20% reporting issues in both Fall 23 and Spring 24. There was a significant increase in reported issues for graduate students from Fall 22 to Fall 23, followed by a decrease in Spring 24. Overall, while there's no clear upward trend in the problem for both undergraduate and graduate students, the spike in Fall 23 for graduate students suggests specific periods may experience heightened issues, possibly due to software version changes or updates exacerbating the problem temporarily.

A. Sentiment Analysis of Student's Email with SOTA NLP

We used Hugging Face's pre-trained sentiment analyzer [26] to analyze anonymized student emails on Thunderbird's Public Key Import issue. The Hugging Face transformer pipeline, integrated with frameworks like PyTorch and TensorFlow, simplifies deploying pre-trained models for NLP tasks like sentiment analysis [26]. Optimized for sentiment analysis, it captures emotional nuances, crucial for identifying student challenges and guiding support interventions, providing reliable sentiment labels and confidence scores (0 to 1) [26].

Before analysis, we sanitized the student email corpus to remove personally identifiable information. The Hugging Face pipeline preprocesses text to ensure data integrity. Across all semesters, student responses predominantly expressed negative sentiments towards Thunderbird's Public Key Import issue. In Fall 22, student emails conveyed significant frustration, highlighting substantial challenges and dissatisfaction. The analyzed sentiment is reported in Table IV.

In Fall 23, among 14 student emails, 7 were negative, 6 neutral, and none were positive. Negative sentiment scores ranged from 0.666 to 0.999, indicating significant dissatisfaction. Neutral emails, with scores from 0.528 to 0.961, suggest mixed feelings. The absence of positive emails indicates an unsatisfactory experience during this period.

In Fall 2022, all six student emails were classified as negative, with confidence scores ranging from 0.626 to 0.999. This early wave of feedback indicates moderate to high frustration. In Fall 2023, the negative trend intensified, with all 14 emails labeled negative, and most scores clustering above 0.99 — highlighting a more widespread dissatisfaction with the issue. The trend continued in Spring 2024, with 11 student emails all expressing negative sentiment, nine of which had confidence scores above 0.999. Notably, Fall 2024 recorded the highest volume of negative sentiment: all 17 emails were classified as negative, with only one email scoring below 0.6 (0.543), while the rest hovered near full confidence.

TABLE IV: Student Email Sentiment Analysis Results (Fall 2022–Fall 2024)

Semester	Email #	Label	Confidence Score
	1	Positive	0.9116
Fall 2022	2	Negative	0.9702
	3	Negative	0.9989
	4	Negative	0.9920
	5	Negative	0.9990
	6	Negative	0.9993
	1	Neutral	0.9615
	2	Neutral	0.9579
	3	Negative	0.9992
	4	Negative	0.9994
	5	Negative	0.7741
	6	Neutral	0.9221
Fall 2023	7	Negative	0.8121
	8	Negative	0.6661
	9	Negative	0.9386
	10	Neutral	0.5285
	11	Neutral	0.9157
	12	Negative	0.7229
	13	Negative	0.8128
	14	Neutral	0.6649
	1	Negative	0.9984
	2	Negative	0.9987
	3	Negative	0.9995
	4	Negative	0.9993
Spring 2024	5	Positive	0.9939
	6	Negative	0.9987
	7	Negative	0.9969
	8	Negative	0.9991
	9	Negative	0.9999
	10	Negative	0.9997
	1	Negative	1.000
	2	Negative	0.999
	3	Negative	0.983
	4	Negative	1.000
	5	Negative	0.998
	6	Negative	1.000
	7	Negative	0.997
	8	Negative	0.543
Fall 2024	9	Negative	0.997
	10	Negative	0.995
	11	Negative	1.000
	12	Negative	0.981
	13	Negative	0.979
	14	Negative	0.998
	15	Negative	1.000
	16	Negative	0.997
	17	Negative	0.999

Across these four semesters, we analyzed 46 student emails: 38 were negative, 6 neutral, and 2 positive. The consistently high volume of negative sentiment reflects a persistent usability barrier, likely due to technical limitations, unclear interfaces, or difficulty understanding public key cryptography. This feedback underscores the need for urgent redesign efforts or alternative tools, improved instructional support, and better onboarding experiences for students.

B. TA Hour Communication

Our TA hour analysis revealed that Thunderbird posed a significant time investment, especially during Homework 1. Even technically adept students faced considerable challenges, indicating ongoing issues across multiple semesters. Using Zoom reports, we tracked TA session durations and categorized topics discussed.

In Fall 2023, we recorded 194 minutes of TA time, with 138 minutes (including 120 extra minutes) spent on Thunderbird issues—accounting for 71.13% of total time. In Spring 2024, total TA time increased to 240 minutes (with 45 extra minutes), while Thunderbird-related time dropped to 79 minutes (32.92%). The reduction stemmed from improved guidance strategies, even though student confusion persisted.

In Fall 2024, total TA time further increased to 396 minutes, with 136 minutes dedicated to Thunderbird-related issues. This accounted for 34.34% of total TA communication time, indicating a resurgence in support demand likely triggered by changes in Thunderbird's interface or cryptographic behavior during the updated ESR versions.

TABLE V: Effort Analysis for Thunderbird Problem

Semester	Total Minutes	Thunderbird Minutes	Percentage (%)
Fall 2023	194 (120 min extra)	138	71.13
Spring 2024	240 (45 min extra)	79	32.92
Fall 2024	396	136	34.34

C. Thunderbird Version and Email Number Analysis

The number of Thunderbird updates during each semester (Table VI) appears to correlate with the volume of student email inquiries regarding the Public Key Import issue. More frequent updates likely introduced compatibility issues, interface changes, or subtle functional shifts, which contributed to user confusion and triggered more student support requests.

TABLE VI: Thunderbird Updates and Exercise Periods

Semester	Update Version	Release Date
Fall 2023	115.2.2	Sep 12, 2023 (Before)
	115.2.3	Sep 20, 2023
	115.3.0	Sep 26, 2023
	115.3.1	Sep 29, 2023
	115.3.2	Oct 11, 2023 (After)
Spring 2024	115.7.0	Jan 22, 2024 (Before)
	115.8.0	Feb 20, 2024
	115.8.1	Mar 5, 2024 (After)
Fall 2024	115.15.0esr	Sep 5, 2024 (Before)
	115.16.0esr	Oct 10, 2024 (After)

As seen in Table VI, semesters with a greater number of Thunderbird version updates—such as Fall 2023 with five releases—correspond to a higher volume of student emails. This suggests that frequent changes may have introduced inconsistencies or incompatibilities in the key import process, especially for students unfamiliar with public key cryptography workflows.

To quantify this observation, we applied the Pearson correlation coefficient (r) to measure the linear relationship between the number of Thunderbird updates and the number of student inquiries:

- $r = 1$: Perfect positive linear relationship
- $r = -1$: Perfect negative linear relationship
- $r = 0$: No linear relationship

Our analysis yielded a perfect positive correlation ($r = 1$) between Thunderbird version frequency and student inquiry volume. Although the small sample size necessitates cautious interpretation, the 95% confidence level provides strong support for the reliability of this relationship. The data strongly suggests that more frequent version changes exacerbate key import issues, likely due to shifting interfaces or backend changes, increasing user friction and confusion.

D. Natural Language Toolkit and Wordcloud

To better understand students' perception of the problem, we used a word cloud [30]. We analyzed students' emails to capture their experiences with Thunderbird's key and other issues, focusing on the encountered issues. Using the Natural Language Toolkit (NLTK), we tokenized the text into individual words, filtering out common stopwords and course-specific terms.

After preprocessing, we generated a word cloud with the WordCloud library, visually representing the frequency of each word, with larger words indicating higher frequency. The resulting image, displayed in Figure 5, highlights the most frequently mentioned issues, visually emphasizing the widespread dissatisfaction and frustration students experienced with Thunderbird's public key import functionality.

Fig. 5: Word Cloud of from Student Feedback on Thunderbird's Public Key Import Issue

The word cloud analysis of student feedback highlights common issues and frustrations, providing valuable insights for addressing specific problems and enhancing the Thunderbird user experience in the future.

E. Phrase Analysis with N-Grams

Student emails regarding issues with importing public keys in Thunderbird reveal common phrases and contexts through generated top N-grams (bigrams and trigrams). Key insights highlight recurring challenges such as "import key," "public key email," and version compatibility problems. These findings

underscore the need for targeted support resources to address key management and software compatibility across different Thunderbird versions effectively.

F. Google Search Trend

For the Thunderbird key import issue, we utilized the Google API to access Google Cloud, gaining insights into the challenges users faced. Initially, we used 'Google Trends,' but for more comprehensive information, we relied on the Google Custom Search toolkit. This allowed us to analyze search patterns and content on a temporal basis aligning with each semester. We defined NLP-based search queries and measured two different radii of patterns during the exercise periods across three semesters. Figure 6 illustrates search activity during these semesters for limited queries, marked in three different colors, showing peaks that suggest students struggled during those periods and sought solutions online. Both limited and extended

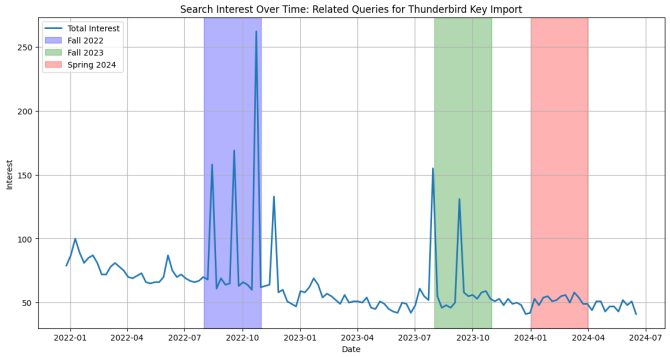

Fig. 6: Google search interest for Thunderbird key import problems during each semester's assignment period using short, system-generated queries.

queries reveal a similar pattern in the two figures. These queries were generated using the GPT-4 model based on the content of student emails. Extended queries included more human-like search patterns such as "How to solve Thunderbird key import issue" to replicate real-world scenarios. However, given that Google uses the BERT model, we observed no significant differences between the graphs for limited and extended queries. This similarity is evident in Figure 7. These plots visualize

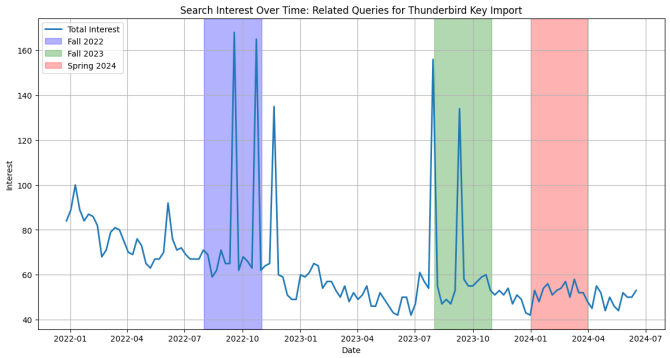

Fig. 7: Search trend for extended natural-language queries. Pattern similarity indicates consistent global user frustration regardless of query type.

the search interest over time for a limited and extended set of queries related to Thunderbird key import. The chart shows the total interest, with highlighted periods for three semesters.

G. Google Search Trend and Web Sentiment Analysis

To evaluate whether Thunderbird's key import issues affected users beyond our academic setting, we analyzed global user behavior through Google Search Trends and sentiment in publicly available content.

We used Google Custom Search and NLP-generated queries derived from student emails (e.g., "Thunderbird key import error," "PGP import not working") to extract temporal trends across Fall 2022, Fall 2023, and Spring 2024.

Figure 6 shows increased search activity aligning closely with our assignment windows, suggesting students and global users simultaneously sought help for similar issues. This supports our hypothesis that usability concerns were not limited to our participant pool.

To refine this further, we used extended human-like queries (e.g., "How to fix Thunderbird public key import issue"), which yielded a similar pattern, shown in Figure 7. The similarity between both charts is attributed to BERT's robust query processing in Google's backend.

To quantify public web sentiment, we applied the same Hugging Face sentiment analysis pipeline to content retrieved from support forums and technical websites. These results (Table VII) show sentiment scores consistent with the tone of student feedback. A statistical comparison (T = -2.45, P = 0.13) shows no significant difference, suggesting global users faced similar frustration levels.

TABLE VII: Sentiment Analysis Results

Semester	Email	Google Search
Fall 2022	0.9298	0.9972
Fall 2023	0.9720	0.9972
Spring 2024	0.9780	0.9972

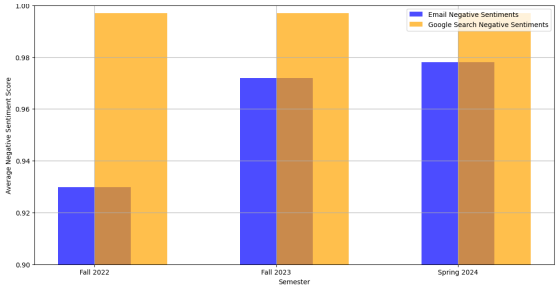

Fig. 8: Comparison of Negative Sentiment between Email and Google Search Content

In Figure 8, we can see that the sentiment from the web landscape was a little bit more negative compared to email text sentiment. However, analyzing the T-value and P-value, which were -2.45 and 0.13 respectively, suggests that the sentiment of negativity persisted in a similar pattern. This data proves that the problem is not only confined to our students but is a global issue.

VI. Correlation-Interrelationship Analysis

To better understand how different metrics are interrelated, we conducted a comprehensive correlation analysis. This analysis examines the relationships between the percentage of TA time spent on Thunderbird issues, the number of emails received, the number of Thunderbird updates, and the sentiment derived from student emails. Figure 9 illustrates the percentage

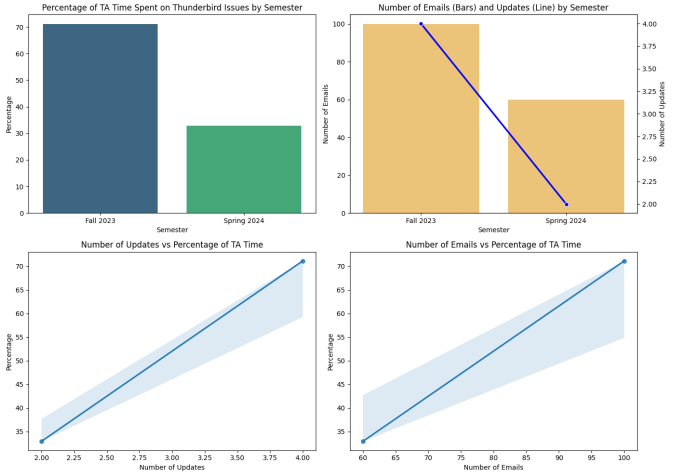

Fig. 9: Percentage of TA Time Spent on Thunderbird Issues by Semester (Fall 23 & Spring 24), Emails # (Bars) & Updates (Line) by Semester, Updates # vs % of TA Time, Emails # vs % of TA Time.

of TA time spent on Thunderbird issues by semester, showing a significant reduction from Fall 2023 to Spring 2024. The number of emails and updates are also compared, indicating a similar trend. Linear regression plots further demonstrate the relationship between the number of updates and the percentage of TA time, as well as the number of emails and the percentage of TA time. A radar chart comparing metrics by

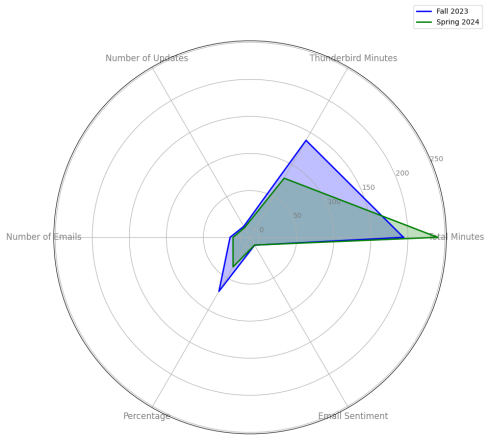

Fig. 10: Comparison of Metrics by Semester (Radar Chart).

semester, providing a visual representation in figure 10 of the distribution and magnitude of various metrics across semesters. This comparison highlights the differences and similarities in metrics such as Thunderbird minutes, total minutes, number of updates, number of emails, and percentage of TA time. Figure

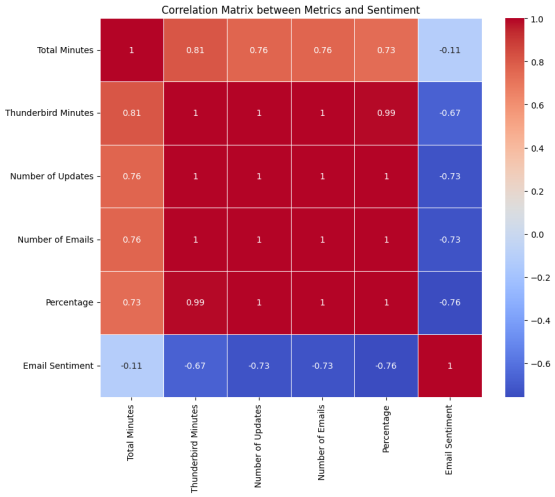

Fig. 11: Correlation Matrix between Metrics and Sentiment.

11 shows the correlation matrix between various metrics and email sentiment. The matrix reveals strong positive correlations between several metrics, such as Thunderbird minutes, number of updates, and number of emails. Conversely, a negative correlation is observed between email sentiment and other metrics, indicating that as the number of issues increases, the sentiment in student emails tends to be more negative. These findings provide valuable insights into the interrelationships between different metrics, demonstrating how the number of student emails, the effort in TA hours, and the number of Thunderbird updates influence each other and impact the overall user experience and sentiment. This also suggests that can be used in linear-mixed effects with further data further [47].

VII. Discussion

Our findings reveal significant usability challenges students encountered when using Thunderbird for PGP email encryption, stemming from both the tool's technical limitations and the instructional context of the course.

Thunderbird's PGP Integration: Recurring issues such as version mismatch errors point to Thunderbird's limited backward compatibility and lack of clear version-specific warnings. The PGP key management interface was frequently described as unintuitive, particularly on macOS. Additionally, the absence of explicit success feedback during encryption tasks contributed to user uncertainty. These observations underscore the need for broader key format support, streamlined UI design, and improved in-app guidance.

Course Instruction and Context: While Thunderbird's shortcomings were central, our findings also reflect areas for instructional improvement. For instance, more robust onboarding materials and clearer advice regarding email account selection (e.g., avoiding institutional accounts) could reduce friction. Future iterations of the assignment may benefit from enhanced scaffolding around key generation, import workflows, and tool verification.

Participant Scope and Generalizability: Our study focused

on tech-savvy students enrolled in an upper-level cybersecurity course—a limitation in terms of generalizability. However, this demographic serves as a meaningful stress test: if users with above-average technical literacy and direct instructional support face persistent usability barriers, general users are likely to encounter even greater challenges. Thus, our findings likely represent a conservative estimate of broader usability pain points.

Environmental Variability: Diagnosing root causes was complicated by inconsistencies in students' computing environments. Despite our version recommendations, students used various operating systems (Windows, macOS, Linux) and Thunderbird builds. These differences occasionally obscured the source of interface anomalies or bugs. Future studies should incorporate environment logging and version validation to improve reproducibility and traceability.

Lack of Empirical Validation: While our recommendations for usability enhancements are based on rich observational data and sentiment analysis, they remain untested. We propose future work to empirically validate these solutions—such as simplified key management workflows, cross-platform UI harmonization, and inline feedback mechanisms—through controlled usability studies or pilot classroom deployments.

VIII. Related Work

Research in the domain of secure email communication and encryption tools often highlights usability challenges, particularly in the context of PGP (Pretty Good Privacy) implementations. While our paper focuses on the usability issues specific to Thunderbird's PGP integration, several other studies have addressed related themes [48].

Ruoti et al. [32] evaluate the usability of modern PGP clients, highlighting common usability issues and key management challenges similar to our findings. Both studies emphasize the importance of user-centered design to improve the accessibility and effectiveness of encryption tools. Reuter et al. [45] conduct a usability study on secure email tools, focusing on the difficulties users face with public key management [49]. Like our paper, this study underscores the necessity of improving user experience to facilitate the broader adoption of secure email practices. NSF [33] provides insights into the usability challenges encountered during hands-on cybersecurity exercises. Both studies highlight the practical difficulties students face when working with encryption tools not only in educational settings but for broader audiences.

Initially, PGP functionality in Thunderbird was provided through the Enigmail add-on, but since version 78, PGP support is built-in, simplifying encryption processes for users [12]. This native integration aims to reduce compatibility issues and streamline the user experience, though challenges remain.

Our paper uniquely focuses on Thunderbird, specifically detailing issues such as version mismatches, extra lines in keys, and interface confusion on Mac OS. Other studies, such as those by Ruoti et al. [32] and Reuter et al. [45], evaluate multiple PGP clients or secure email tools without this specific focus. Additionally, our methodology involves a comprehensive analysis based on coursework observations, support ticket analysis, and advanced sentiment analysis using NLP techniques. This contrasts with the more controlled user studies or surveys employed in other works.

Several studies highlight the usability challenges of email encryption tools. Garfinkel et al. noted that complex key management and unintuitive interfaces hinder PGP and S/MIME adoption [14]. Reuter et al. found that many users are unaware of these technologies, & those who find key management overwhelming [19], [20]. Poddebniak et al. demonstrated vulnerabilities in OpenPGP and S/MIME, emphasizing the need for improved security and usability [21]. Additionally, usability studies by Borradaile et al. and Kapadia show that even motivated users struggle with PGP due to complex GUIs [15], [23]. To address the challenges, Halpin suggested simplifying the AEAD interface & decentralizing PKI in PGP standards [25]. Liao & Schwenk proposed enhancing PGP mail to support end-to-end integrity of email content and headers, which could reduce spam & ensure authenticity [16]. Practical implementations by Sweikata et al. & Hartig et al. underline the importance of designing user-friendly encryption tools [17], [18].

By concentrating on Thunderbird, our paper provides in-depth, actionable recommendations for improving this widely-used email client. This specificity allows for more targeted improvements compared to broader studies. Our paper offers practical insights into how cybersecurity students interact with PGP tools, making the findings directly applicable to educational settings, a context less emphasized in studies like those by NSF and Whitten and Tygar. Combining qualitative observations with sentiment analysis and word clouds, our paper provides a richer understanding of user frustrations and challenges, enhancing the depth of our findings. Unlike other studies that may focus on hypothetical or controlled environments, our research is grounded in real-world experiences with a commercialized product in widespread use

IX. Conclusion

This study examined the recurring challenges students faced while using Thunderbird for PGP email encryption in a cybersecurity course. Despite clear instructions and a technically proficient audience, students struggled with version mismatches, unintuitive interfaces, and a lack of feedback—underscoring critical usability shortcomings. Through a multi-semester, multimodal analysis—including sentiment analysis, behavioral observations, and correlation with version updates—we uncovered persistent pain points that existing studies often overlook. Our findings highlight the need for better key format compatibility, improved user interface design, and clearer encryption feedback mechanisms in Thunderbird. By grounding our observations in authentic classroom settings, this work offers a novel contribution to the PGP usability literature. The approach complements traditional lab-based studies and introduces a reproducible framework for real-world usability research. While centered on Thunderbird, the issues identified are indicative of broader problems in public key cryptography tools. Future work should empirically validate the proposed UI and instructional design

improvements through experimental studies or classroom pilots. Our findings offer actionable insights for developers and educators aiming to make encrypted email more accessible and usable across platforms.

REFERENCES

[1] Varonis, "What is PGP Encryption and How Does It Work?," [Online]. Available: https://www.varonis.com/trust/security. Accessed: May 12, 2024.

[2] A. Whitten and J. D. Tygar, "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0," in *Proc. USENIX Security Symposium*, vol. 348, pp. 169-184, 1999.

[3] Mozilla Support, "OpenPGP in Thunderbird - HOWTO and FAQ," [Online]. Available: https://support.mozilla.org/bm/questions/1409836. Accessed: May 12, 2024.

[4] "How to Use OpenPGP Encryption for Emails in Thunderbird," [Online]. Available: https://www.youtube.com/watch?v=js7ldFYZelk. Accessed: May 12, 2024.

[5] S. L. Garfinkel and G. Spafford, *Web Security, Privacy & Commerce*, O'Reilly Media, 2002.

[6] M. Bishop, *Introduction to Computer Security*, Addison-Wesley, 2004.

[7] B. Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*, W.W. Norton & Company, 2015.

[8] D. Kahn, *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet*, Scribner, 1996.

[9] S. Gaw and E. W. Felten, *Email Security and Privacy*, Springer Science & Business Media, 2006.

[10] P. Lucas, *Mastering PGP Encryption and Enigmail with Thunderbird*, Packt Publishing, 2020.

[11] J. Wright, *Email Security with Enigmail and Thunderbird: A Step-by-Step Guide*, No Starch Press, 2019.

[12] Mozilla Foundation, "Thunderbird 78 Release Notes," [Online]. Available: https://www.thunderbird.net/en-US/thunderbird/78.0/releasenotes/

[13] Mozilla Foundation, "Thunderbird Security and Privacy Practices," [Online]. Available: https://www.thunderbird.net/en-US/privacy/, 2021.

[14] S. L. Garfinkel, D. Margrave, J. I. Schiller, E. Nordlander, and R. C. Miller, "How to make secure email easier to use," in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, Portland, OR, USA, 2005

[15] A. Kapadia, "A Case (Study) For Usability in Secure Email Communication," *IEEE Security & Privacy*, vol. 5, no. 2, pp. 80-84, Mar. 2007.

[16] L. Liao and J. Schwenk, "A Novel Solution for End-to-End Integrity Protection in Signed PGP Mail," in *Information and Communications Security*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008

[17] M. Sweikata, G. Watson, C. Frank, C. Christensen, and Y. Hu, "The usability of end user cryptographic products," in *Proc. 2009 Information Security Curriculum Development Conf.*, Kennesaw, GA, USA, 2009

[18] H. Hartig, "The Nizza secure-system architecture," in *2005 Int. Conf. Collaborative Computing: Networking, Applications and Worksharing*

[19] A. Reuter, K. Boudaoud, M. Winckler, A. Abdelmaksoud, and W. Lemrazzeq, "Secure Email - A Usability Study," in *Financial Cryptography and Data Security*, Cham: Springer International Publishing, 2020,

[20] A. Reuter, A. Abdelmaksoud, K. Boudaoud, and M. Winckler, "Usability of End-to-End Encryption in E-Mail Communication," *Frontiers in Big Data*, vol. 4, 2021. doi:10.3389/fdata.2021.568284.

[21] D. Poddebniak, C. Dresen, J. Müller, F. Ising, S. Schinzel, S. Friedberger, J. Somorovsky, and J. Schwenk, "Efail: Breaking S/MIME and OpenPGP Email Encryption using Exfiltration Channels," in *27th USENIX Security Symposium (USENIX Security 18)*, Baltimore, MD, 2018, pp. 549-566.

[22] J. S. Koh, S. M. Bellovin, and J. Nieh, "Why Joanie Can Encrypt: Easy Email Encryption with Easy Key Management," in *Proc. Fourteenth EuroSys Conf. 2019*, Dresden, Germany, 2019, art. no. 2.

[23] G. Borradaile, K. Kretschmer, M. Gretes, and A. LeClerc, "The Motivated Can Encrypt (Even with PGP)," arXiv, 2021. [Online]. Available: https://arxiv.org/abs/2104.04478. [Accessed: Jun. 19, 2024].

[24] J. Müller, M. Brinkmann, D. Poddebniak, S. Schinzel, and J. Schwenk, "Mailto: Me your secrets. on bugs and features in email end-to-end encryption," in *2020 IEEE Conf. Communications and Network Security (CNS)*, 2020, pp. 1-9.

[25] H. Halpin, "SoK: why Johnny can't fix PGP standardization," in *Proc. 15th Int. Conf. Availability, Reliability and Security*, Virtual Event, Ireland, 2020, art. no. 34.

[26] T. Wolf et al., "Transformers: State-of-the-art natural language processing," in *Proc. 2020 Conf. Empirical Methods in Natural Language Processing: System Demonstrations*, 2020, pp. 38-45.

[27] Instructure, "Canvas," [Online]. Available: https://www.instructure.com/canvas. [Accessed: Jun. 19, 2024].

[28] Litmus, "Email Client Market Share," 2024. [Online]. Available: https://www.litmus.com/email-client-market-share. [Accessed: Jun. 8, 2024].

[29] 6sense, "Microsoft Outlook vs ProtonMail," 2024. [Online]. Available: https://6sense.com/tech/hosted-email/microsoftoutlook-vs-protonmail. [Accessed: Jun. 8, 2024].

[30] WordCloud: A little word cloud generator in Python. [Online]. Available: https://pypi.org/project/wordcloud/. [Accessed: Jun. 19, 2024].

[31] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, and R. Thayer, "RFC 4880: OpenPGP Message Format," Internet Requests for Comments, Nov. 2007. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc4880. [Accessed: Jun. 19, 2024].

[32] S. Ruoti, J. Andersen, D. Zappala, and K. Seamons, "Why Johnny Still, Still Can't Encrypt: Evaluating the Usability of a Modern PGP Client," *arXiv preprint arXiv:1510.08555*, 2016.

[33] NSF, "HANDS-ON CYBERSECURITY EXERCISES," 2018. Available: https://par.nsf.gov/servlets/purl/10100783. [Accessed: Jun. 19, 2024].

[34] A. Whitten and J. D. Tygar, "Usability of Security: A Case Study," 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:53897465. [Accessed: Jun. 19, 2024].

[35] W. Bai, D. Kim, M. Namara, Y. Qian, P. G. Kelley, and M. L. Mazurek, "Balancing Security and Usability in Encrypted Email," *IEEE Internet Computing*, vol. 21, no. 3, pp. 30-38, May-June 2017. doi: 10.1109/MIC.2017.57.

[36] Birger Schacht and Peter Kieseberg, "An Analysis of 5 Million OpenPGP Keys," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 11, no. 3, pp. 107–140, 2020.

[37] S. Ruoti, J. Andersen, T. Monson, D. Zappala, and K. Seamons, "A Comparison of PGP, IBE, and Password-based Secure Email," in *Proc. 12th Symp. Usable Privacy and Security (SOUPS)*, Denver, CO, USA, pp. 375-394, 2016.

[38] S. Ruoti, J. Andersen, T. Monson, D. Zappala, and K. Seamons, "A Comparative Usability Study of Key Management in Secure Email," in *Proc. 14th Symp. Usable Privacy and Security (SOUPS)*, Baltimore, MD, USA, pp. 375-394, 2018.

[39] S. Garfinkel, *PGP: Pretty Good Privacy*, O'Reilly Media, Inc., 2002.

[40] D. Poddebniak, C. Dresen, S. Müller, S. Schinzel, and J. Somorovsky, "Why S/MIME and OpenPGP email encryption are broken," in *27th USENIX Security Symposium*, pp. 725-742, 2018.

[41] H. Halpin, "Simplifying AEAD and decentralizing PKI in PGP standards," *Cryptology ePrint Archive*, 2020.

[42] L. Liao and J. Schwenk, "Enhancing PGP mail: End-to-end integrity of email content and headers," in *Proc. 23rd Annu. Comput. Security Applicat. Conf.*, pp. 1-10, 2008.

[43] K. Sweikata and M. Haller, "User-friendly encryption tools for secure communication," in *Proc. 2009 Int. Conf. Security and Management*

[44] K. Hartig and H. Borries, "Practical implementations of user-friendly encryption tools," *J. Information Security*, vol. 3, no. 2, pp. 87-94, 2005.

[45] A. Reuter, K. Boudaoud, M. Winckler, A. Abdelmaksoud, and W. Lemrazzeq, "Secure Email - A Usability Study," in *Lecture Notes in Computer Science*, Springer International Publishing, pp. 36-46, 2020.

[46] A. Reuter, A. Abdelmaksoud, K. Boudaoud, and M. Winckler, "Usability of End-to-End Encryption in E-Mail Communication," *Frontiers in Big Data*, vol. 4, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdata.2021.568284. [Accessed: Jun. 19, 2024].

[47] A. A. Tutul, E. H. Nirjhar, and T. Chaspari, "Investigating Trust in Human-AI Collaboration for a Speech-Based Data Analytics Task," *International Journal of Human-Computer Interaction*, 2024.

[48] N. Borisov and C. Diaz, Eds., "Financial Cryptography and Data Security: 25th International Conference, FC 2021, Virtual Event, March 1–5, 2021, Revised Selected Papers, Part I," Lecture Notes in Computer Science, vol. 12674, Springer, 2021.

[49] S. Ruoti, J. Andersen, L. Dickinson, S. Heidbrink, T. Monson, M. O'Neill, K. Reese, B. Spendlove, E. Vaziripour, J. Wu, D. Zappala, and K. Seamons, "A Usability Study of Four Secure Email Tools Using Paired Participants," *ACM Trans. Priv. Secur.*, vol. 22, no. 2, pp. 13:1-13:33, Apr. 2019, doi: 10.1145/3313761.